A new method to learn growth curves of beef cattle using a factorization approach

Jaime Alonso*, Jorge Díez*, Oscar Luaces*, Antonio Bahamonde*

Artificial Intelligence Center
University of Oviedo at Gijón, Asturias, Spain

Abstract

The evolution of cattle weight is a very important issue for beef cattle breeders. The weights of the animals of a herd are usually available at different ages and it is intended to predict the trajectory that will follow the weight of each animal. In this paper, we address this problem as a Recommender System. In this case, the users would be the animals, and the items would be the ages of weight measurements. The values of the items would be the measured weights at a given age. As in Recommender Systems the aim is to complete the valuation matrix (weights) in an individualized way (that is, adapted to each animal). A matrix factorization system is devised to learn weights using all the available characteristics of the animals. The weights thus obtained are compared with a linear regression that adequately estimates the general evolution of the herd, but not the individual evolution of each animal. To illustrate the benefits of this approach, we used a real world dataset of cattle of the breed Avileña-Negra Ibérica and crossbreeding with sires of Charolais and Limousin.

Keywords: Recommender systems, Matrix factorization, Stochastic gradient descent, Growth curves, Beef cattle

---

*Corresponding author: Tel: +34 985 182 032
Email addresses: `jalonso@uniovi.es` (Jaime Alonso), `jdiez@uniovi.es` (Jorge Díez), `oluaces@uniovi.es` (Oscar Luaces), `abahamonde@uniovi.es` (Antonio Bahamonde)

1. Introduction

Live weight of beef cattle is the main feature affecting carcass performance and hence the incomes of breeders. Thus, the estimation of the weight of bovines as a function of time is very important for meat producers. Predicting growth can help breeders to decide the best time to slaughter in order to increase their economic benefits.

There are different approaches to study weight evolutions in the literature. In (West et al., 2001) they propose a universal curve that describes the growth of many diverse species, transforming time and weight data to a common dimensionless scale. Live weight of bovines is often estimated from easily accessible morphometric characteristics (Enevoldsen and Kristensen, 1997; Coopman et al., 2009). Sometimes, to obtain the body measurements the authors use digital image processing (Stajnko et al., 2008; Tasdemir et al., 2011). In other cases, genetic information about animals it is also considered. A comparison of different genetics methods can be seen in (Jaffrézic and Pletcher, 2000). These methods try to obtain the capability of animals to transmit genetically the gain of weight to their progeny (Arango et al., 2004; Freetly et al., 2011).

In this paper, we use Machine Learning procedures to learn a model to predict weights of animals of *Avileña-Negra Ibérica*. We want to anticipate the weight of each single animal (Díez et al., 2003; Alonso et al., 2007, 2013, 2015). *Avileña-Negra Ibérica* is a beef breed of central Spain. Their carcasses are characterized by an intermediate muscle conformation (69.1%) and fatness level (12.6%) (Albertí et al., 2008). The market target of these carcasses is made up of those consumers that prefer tender meat but with an intense flavor (Díez et al., 2006). Due to calving ease, straightbred cows of *Avileña-Negra Ibérica* are used in crossbreeding programs with sires of *Charolais* and *Limousin* breeds.

The approach presented here is based on Matrix Factorization (MF). In recent years, MF techniques for machine learning have been attracting more and more attention, especially since a recommender system based on MF won the *Netflix* prize (Koren et al., 2009). MF algorithms were used in many application

2

fields such as medical recommendations (Zhang et al., 2016), audience selection (Kanagal et al., 2013), text analysis (Ji and Eisenstein, 2014), or evaluating open-response assignments in Massive Open Online Courses (MOOCs) (Luaces et al., 2015a).

Dimensionality reduction is another important task of machine learning where MF is widely used. In this context, the purpose is to obtain a simplified lower-rank approximation of the original dataset. This allows presenting, possibly geometrically, the structure that may be inherent in a dataset, mainly the relationship between objects and their attributes. Dimensionality reduction with MF has been successfully applied in feature selection (Wang et al., 2015), image analysis (Khurana et al., 2015; Moon et al., 2016), to identifying user preferences (Luaces et al., 2015b) and in many other problems along with clustering techniques.

We took advantage of matrix factorization techniques in two different ways. On the one hand, we develop an algorithm capable of predicting the evolution of animal weights with good accuracy. On the other hand, our system projects the information of animals in a Euclidean space with a reduced dimension. This embedding enables visualization and clustering of animals based on some trait, such as its growth potential.

Data in our problem are represented by means of a matrix in the same way as recommender systems do. In such systems, ratings of users about a kind of products are arranged in matrices. Users are laid out in rows, while products are set out in columns. Users only rate a small subset of the product database. Therefore, even the most popular items have very few ratings; thus the matrix has very low density. In our case, the entries of the matrix are the live weights of animals, the rows contain the information of each animal and the columns contain the information on the animal's age when the corresponding weight was taken. Since animals have been weighed only a few times at different ages we get a sparse matrix. We have to learn a function able to fill in the gaps of the matrix (weights missing at certain ages) with the least possible error. Our data would be similar to those shown in Table 1.

Table 1: Representation of the data of our problem in matrix form.

	Age_1	Age_2	Age_3	Age_4	Age_5
Animal_1		220	290	340	
Animal_2	180		275		400
Animal_3		150	196	230	
Animal_4	210	258		389	460

Both animals and ages are represented by feature vectors. These vectorial descriptions are essential for the proper performance of the system. They allow to establish differences between animals based on features such as sex, breed, body condition or birthing season. Moreover, descriptions can be enriched if needed with some new features of animals or ages.

It is important to emphasize that our method allows to obtain weight trajectories adapted to the specific data of each animal. Other systems, such as regression, only differentiate weight trajectories of the animals in a constant (specific to each animal). As a result, all the trajectories are parallel. This simplification implies the assumption that the weight of all animals increases in the same way, which is unacceptable.

2. Material and methods

2.1. Data

The dataset used in this study was obtained from several feedlots of *Avileña-Negra Ibérica* breed of cattle. This set contains two types of animals: purebred and crossbred animals. Crossbred animals are always from *Avileña-Negra Ibérica* mother and father from Charolais or Limousin breed. Information on each animal consists of its identification, origin, date of birth and live weight measured at different ages. Only animals that have 4 or more measurements were used because a significant sequence of weights is required in order to *learn* the way animals grow. Cattle are transferred to a feedlot at ages that range

4

Table 2: Description of the number of animals and weights divided by sex and father's breed

Abbreviation	Father's breed	Sex	#Animals	#Weights
Avi_M	Avileña	male	6353	32317
Avi_F	Avileña	female	347	1730
Cha_M	Charolais	male	1706	8258
Cha_F	Charolais	female	1670	7741
Lim_M	Limousin	male	481	2360
Lim_F	Limousin	female	435	2072
			10992	54478

from 90 to 365 days. The number of animals and weights to work with are detailed in Table 2.

We would like to emphasize that available data on animals are quite diverse in several aspects:

- There are important differences in growth by sex and breed.

- Animals enter the feedlot at quite different ages.

- The ratio of weight to age varies considerably when cattle enter the feedlot.

- Growth rates of animals are very different even if they have the same sex and breed.

This disparity complicates the learning task considerably. Figure 1 shows weight curves of some animals to illustrate differences among them. It has been necessary to design a method capable to properly reflect all this variability.

2.2. Learning Method

Let $\mathcal{A}$ be a set of *animals* and let $\mathcal{D}$ be a set of *ages* in which animals are weighted. Let us consider a partially defined matrix

$$\mathcal{M} = (m_{ij} \; : \; i \in \mathcal{A}, \, j \in \mathcal{D}) \tag{1}$$
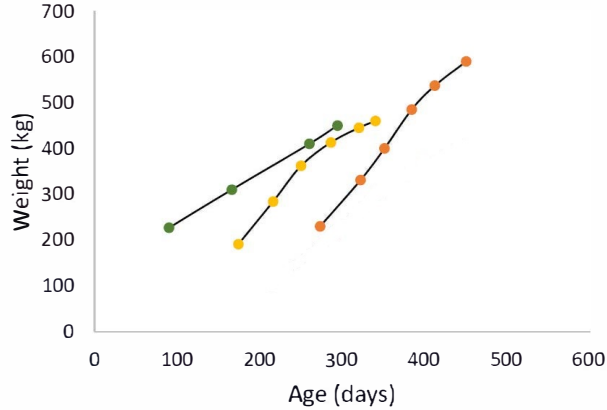
5

Figure 1: Sample growth trajectories of animals showing differences in shape, slope and initial age (the age when animals enter the feedlot)

where m_{ij}, if available, is the live weight (in kilograms) of the animal i at the age j. Both animals and ages will be represented by vectors of features. These vectors are used to describe specific characteristics of the animals/ages. We explain this representation in more detail in Section 2.3.

Our main objective is to learn the weights of animals making use of matrix completion techniques. We look for a function

$$g : \mathcal{A} \times \mathcal{D} \rightarrow \mathbb{R} \tag{2}$$

to compute weights of animals at any age. In other words, we will learn a function g with the following general expression:

$$g(\boldsymbol{a}, \boldsymbol{d}) = \sum_{i,j} \alpha_{ij} \, \boldsymbol{a_i} \, \boldsymbol{d_j}. \tag{3}$$

This function is just a weighted sum of the products of the vectors representing animals and ages. So, the weight of an animal at a given age depends on all possible combinations of descriptors of animal i ($\boldsymbol{a_i}$) and descriptors of age j ($\boldsymbol{d_j}$). But this expression may have too many parameters to learn, the

components of a matrix $\mathbf{\Lambda} = (\alpha_{ij} : i, j)$. Thus, we search for a couple of matrices such that $\mathbf{\Lambda}$ can be *factorized* through them.

$$g(\boldsymbol{a}, \boldsymbol{d}) = \sum_{i,j} \alpha_{ij} \, \boldsymbol{a_i} \, \boldsymbol{d_j} = \boldsymbol{a}^T \mathbf{\Lambda} \, \boldsymbol{d} = \boldsymbol{a}^T \boldsymbol{W}^T \boldsymbol{V} \boldsymbol{d} = \langle \boldsymbol{W} \boldsymbol{a}, \boldsymbol{V} \boldsymbol{d} \rangle. \qquad (4)$$

Therefore, the parameters to be learned have a clear geometrical interpretation. They are mappings (embedding) from the spaces of animals and ages into a common Euclidean space, $\mathbb{R}^k$, for some integer k.

$$\mathcal{A} \to \mathbb{R}^k, \quad \boldsymbol{a} \mapsto \boldsymbol{W} \boldsymbol{a}$$
$$\mathcal{D} \to \mathbb{R}^k, \quad \boldsymbol{d} \mapsto \boldsymbol{V} \boldsymbol{d}. \qquad (5)$$

This common space summarizes all the information of each object (animal or age) in a vector of k components. If k is 2 or 3, then those objects can be represented graphically. This representation on the Euclidean space allows clusters of similar objects, not in their initial characteristics but in their behavior in the learning task being studied.

Finally, we add a component with constant value 1 to the vector descriptions of animals and ages, in order to make a more general expression of (3):

$$\boldsymbol{a} \leftarrow \begin{bmatrix} \boldsymbol{a} & 1 \end{bmatrix},$$
$$\boldsymbol{d} \leftarrow \begin{bmatrix} \boldsymbol{d} & 1 \end{bmatrix}. \qquad (6)$$

Then, the prediction function of the weights is as follows:

$$g(\boldsymbol{a}, \boldsymbol{d}) = \langle \boldsymbol{W} \boldsymbol{a}, \boldsymbol{V} \boldsymbol{d} \rangle$$
$$= \sum_{i,\, j} \alpha_{ij} \, \boldsymbol{a_i} \, \boldsymbol{d_j} + \sum_i \beta_i \, \boldsymbol{a_i} + \sum_j \delta_j \, \boldsymbol{d_j} + \mu. \qquad (7)$$

where μ is the intercept term, β_i and δ_j are the coefficients of the variables of animals and ages, respectively.

Linear regression can also be applied to estimate the weight of the animals. In this case, the regression function could be written as:

$$\text{reg}(\boldsymbol{a}, \boldsymbol{d}) = \sum_i \tilde{\beta}_i \, \boldsymbol{a_i} + \sum_j \tilde{\delta}_j \, \boldsymbol{d_j} + \tilde{\mu}. \qquad (8)$$

If we compare the equations (7) and (8), we appreciate that the difference is in the sum of the cross products of vectors representing animals and ages. This makes the regression expression easier to learn than the factorization one because it has fewer parameters. However, the regression formula only differentiates the trajectories of the evolutions of the weights in a constant value (specific to each animal). Therefore, all the trajectories will be parallel. In contrast, our formula allows to obtain different trajectories adapted to the features of each animal.

In (7) we look for matrices that minimize the mean square error. Thus, for

$$S = \{(\boldsymbol{a}, \boldsymbol{d}) : \mathcal{M}(\boldsymbol{a}, \boldsymbol{d}) \text{ is defined}\} \tag{9}$$

where $\mathcal{M}$ was defined in (1), the error is calculated as:

$$\text{err}(g, \mathcal{M}) = \frac{1}{2} \sum_{(a,d) \in S} \left(\langle \boldsymbol{W}\boldsymbol{a}, \boldsymbol{V}\boldsymbol{d} \rangle - m_{ad} \right)^2. \tag{10}$$

Then we will look for the matrices that solve the following convex optimization problem:

$$(\boldsymbol{W}^*, \boldsymbol{V}^*) = \underset{\boldsymbol{W}, \boldsymbol{V}}{\text{argmin}} \ \left(\text{err}(g, \mathcal{M}) + \nu \ \text{r}(\boldsymbol{W}, \boldsymbol{V}) \right). \tag{11}$$

That is, we look for matrices that minimize the mean square error and a regularization term. In our case, the square of the Frobenius norm was adopted as the regularization term:

$$\text{r}(\boldsymbol{W}, \boldsymbol{V}) = \left\| \boldsymbol{W} \right\|_F^2 + \left\| \boldsymbol{V} \right\|_F^2. \tag{12}$$

To solve the optimization problem we use a Stochastic Gradient Descent (SGD) algorithm (Bottou, 2010). This method iteratively updates the parameters to learn until a stopping criterion is reached. The parameters of the model are updated as follows:

$$\theta \leftarrow \theta - \gamma \left(\frac{\partial err}{\partial \theta} + \nu \frac{\partial r}{\partial \theta} \right). \tag{13}$$

In our case, since the parameters we want to learn are the matrices $\boldsymbol{W}$ and $\boldsymbol{V}$, the above expression can be written as:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} - \gamma \left(\frac{\partial err}{\partial \boldsymbol{W}} + \nu \frac{\partial r}{\partial \boldsymbol{W}} \right)$$
$$\boldsymbol{V} \leftarrow \boldsymbol{V} - \gamma \left(\frac{\partial err}{\partial \boldsymbol{V}} + \nu \frac{\partial r}{\partial \boldsymbol{V}} \right). \tag{14}$$

The partial derivatives of the error function are computed as follows:

$$\frac{\partial err}{\partial \boldsymbol{W}} = \frac{\partial}{\partial \boldsymbol{W}} \frac{1}{2} (\langle \boldsymbol{W}\,\boldsymbol{a}, \boldsymbol{V}\,\boldsymbol{d} \rangle - m_{ad})^2$$
$$= \frac{\partial}{\partial \boldsymbol{W}} \frac{1}{2} (\langle \boldsymbol{W}\,\boldsymbol{a}, \boldsymbol{V}\,\boldsymbol{d} \rangle^2 + m_{ad}{}^2 - 2\,m_{ad}\langle \boldsymbol{W}\,\boldsymbol{a}, \boldsymbol{V}\,\boldsymbol{d} \rangle) \tag{15}$$

On one side, we apply:

$$\frac{\partial(\langle \boldsymbol{W}\,\boldsymbol{a}, \boldsymbol{V}\,\boldsymbol{d} \rangle^2)}{\partial \boldsymbol{W}} = \frac{\partial(\boldsymbol{a}^T \boldsymbol{W}^{\,T}(\boldsymbol{V}\,\boldsymbol{d})(\boldsymbol{V}\,\boldsymbol{d})^T \boldsymbol{W}\,\boldsymbol{a})}{\partial \boldsymbol{W}} = 2\,(\boldsymbol{V}\,\boldsymbol{d})(\boldsymbol{V}\,\boldsymbol{d})^T \boldsymbol{W}\,\boldsymbol{a}\,\boldsymbol{a}^T$$
$$\tag{16}$$

In addition, for an example $(\boldsymbol{a}, \boldsymbol{d}) \in S$:

$$\frac{\partial(2\,m_{ad}\langle \boldsymbol{W}\,\boldsymbol{a}, \boldsymbol{V}\,\boldsymbol{d} \rangle)}{\partial \boldsymbol{W}} = 2\,m_{ad} \frac{\partial(\boldsymbol{a}^T \boldsymbol{W}^{\,T} \boldsymbol{V}\,\boldsymbol{d})}{\partial \boldsymbol{W}} = 2\,m_{ad} \boldsymbol{V}\,\boldsymbol{d}\,\boldsymbol{a}^T \tag{17}$$

We obtain the final expression of the partial derivative of the error function with respect to $\boldsymbol{W}$ using (16) and (17) in (15):

$$\frac{\partial\,\mathrm{err}}{\partial \boldsymbol{W}} = (\boldsymbol{V}\,\boldsymbol{d})(\boldsymbol{V}\,\boldsymbol{d})^T \boldsymbol{W}\,\boldsymbol{a}\,\boldsymbol{a}^T - m_{ad}\boldsymbol{V}\,\boldsymbol{d}\,\boldsymbol{a}^T \tag{18}$$

Analogously, the expression of the partial derivative of the error function with respect to $\boldsymbol{V}$ is given by:

$$\frac{\partial\,\mathrm{err}}{\partial \boldsymbol{V}} = (\boldsymbol{W}\,\boldsymbol{a})(\boldsymbol{W}\,\boldsymbol{a})^T \boldsymbol{V}\,\boldsymbol{d}\,\boldsymbol{d}^T - m_{ad}\boldsymbol{W}\,\boldsymbol{a}\,\boldsymbol{d}^T \tag{19}$$

Finally, the regularization derivatives are:

$$\frac{\partial\,\mathrm{r}}{\partial \boldsymbol{W}} = 2\,\boldsymbol{W} \quad , \quad \frac{\partial\,\mathrm{r}}{\partial \boldsymbol{V}} = 2\,\boldsymbol{V}. \tag{20}$$

2.3. Feature vectors of animals and ages

The vectorial descriptions of the animals are very important because they allow to establish specific differences among them. Animals are described with a set of features, both measurable and non-measurable. All this information is used by our system to improve the accuracy.

Let us remember that a set of live weights is available for each animal $a \in \mathcal{A}$ taken at different ages:

$$a : ((ag_1, lw_1), (ag_2, lw_2), \ldots, (ag_n, lw_n)) \tag{21}$$

where ag_i is the i-th age (in days) and lw_i is the i-th live weight (in kilograms) for animal a.

We use a set of variables to physically describe animals and to reflect their conditions when they get at feedlot (as we explained in Section 2.1). It is important to note that some of them are based on the m first measurements of each animal. The goal is to obtain models that improve accuracy as the number of animal measurements increase. These are the variables:

- ag_1: age when animal enters the feedlot.

- ag_m: m-th age.

- $\overline{ag}_m$: average age $(ag_1..ag_m)$.

- lw_1: live weight when animal enters the feedlot.

- lw_m: m-th live weight.

- $\overline{lw}_m$: average live weight $(lw_1..lw_m)$.

- $rwa_1 = \frac{lw_1}{ag_1}$: ratio of weight to age when animal enters the feedlot.

- $adg_m = \frac{lw_m - lw_1}{ag_m - ag_1}$: average daily gain considering the m first weights.

Moreover, to characterize the animals we have also used some categorical features:

Table 3: Statistical information about data of animals (ag_1, lw_1: age in days and live weight in Kg. when animals come in at feedlot; Avi: avileña, Cha: charolais, Lim: limousin)

	ag_1	lw_1	lw_1/ag_1	Age	Weight
Mean	220.99	238.34	1.10	330.87	380.42
Std.	45.90	52.05	0.23	93.09	115.00
Min.	90	66	0.5	90	66
Max.	339	441	3.03	674	804

	Sex		Father's breed		
	Male	Female	Avi	Cha	Lim
#Weights	42935	11543	34047	15999	4432
Percentage	79%	21%	63%	29%	8%

- Sex: male or female.

- Father's breed: avileña, charolais or limousin.

Put all together, the vectorial representation of an animal $a \in \mathcal{A}$ can be written as:

$$\boldsymbol{a} = (ag_1, ag_m, \overline{ag}_m, lw_1, lw_m, \overline{lw}_m, rwa_1, adg_m, \text{sex}, \text{father's\_breed}) \qquad (22)$$

In turn, the vectorial description for an age $d \in \mathcal{D}$ is:

$$\boldsymbol{d} = (ag, ag^2, ag^3) \qquad (23)$$

where the feature vector is formed by the age (in days) where the corresponding weight was measured as well as the square and cube of such value. The purpose of including these values is to allow our method to learn growth trajectories based on a non-linear relation with the age. Table 3 shows descriptive statistics on the data used in our system to describe animals and ages.

3. Experimental results

In this section we report the results of a set of experiments designed to evaluate the approach proposed in this paper. First we introduce the datasets

used in the experiments. Then we describe the experimental procedure and how to compute the accuracy of the different models. Finally, we discuss the results and we depict some graphs to illustrate the visualization possibilities of the factorization approach.

3.1. Datasets

The datasets consist of examples drawn from the information about the animals. An animal $a \in \mathcal{A}$ with n measurements of live weight and age expressed as $((ag_1, lw_1), (ag_2, lw_2), \ldots, (ag_n, lw_n))$ yields n examples of this form:

$$(\boldsymbol{a}, \boldsymbol{d_1}, lw_1), (\boldsymbol{a}, \boldsymbol{d_2}, lw_2), \ldots, (\boldsymbol{a}, \boldsymbol{d_n}, lw_n) \tag{24}$$

where $\boldsymbol{a}$ is the feature vector of animal a, $\boldsymbol{d_i}$ is the feature vector of age ag_i and lw_i is the live weight of the animal at age ag_i. The targets are the weights at each age.

Numerical variables have been *standardized* so that the value of the variable minus the mean is divided by the standard deviation. Non-numerical features have been *binarized*. That is, each original feature is replaced as many binary variables as the possible values it can take. A binary variable can only take values 0/1 meaning absence/presence of the trait it represents. For example, the *sex* feature is replaced by two binary variables: *male* and *female*. Then, the value of *male* will be 1 for a male animal and 0 if the animal is female.

Let us take a detailed look at the structure of the examples of a given pure-bred male of Avileña. We have 6 measures, of the form (age, weight), taken on this animal: $((273, 230), (322, 331), (351, 400), (384, 485), (412, 537), (420, 517))$. Table 4 shows the portion of the examples that corresponds to the vectorial description of the animal, with m ranging from 1 to 4. The last pair, $(420, 517)$, will be used as the target to be predicted so we will represent the age, 420, as indicated in (23), i.e. $(420, 420^2, 420^3)$, for all the descriptions in Table 4.

3.2. Experiments: description and settings

We have conducted a number of experiments to evaluate the influence of some important issues of the data:

Table 4: Feature vectors of a purebred male of Avileña for m in range from 1 to 4. Animal measures, in the form (age, weight), are the followings: $((273, 230), (322, 331), (351, 400), (384, 485), (412, 537), (420, 517))$. Columns with (*) are not used for $m=1$

			Feature vector of given animal										
m	ag_1	ag_m	$\overline{ag}_m$	lw_1	lw_m	$\overline{lw}_m$	rwa_1	adg_m	Male	Female	Avi	Cha	Lim
1	273	273^*	273^*	230	230^*	230^*	0.84	nan^*	1	0	1	0	0
2	273	322	297.5	230	331	280.5	0.84	2.06	1	0	1	0	0
3	273	351	312	230	400	315	0.84	2.18	1	0	1	0	0
4	273	384	328.5	230	485	357.5	0.84	2.29	1	0	1	0	0

1. The type of animals: all together or separated by breed and sex. Our system has been tested with datasets made up of all animals and with datasets formed by animals divided by sex and breed. Obviously, in the last case we do not use the variables representing sex and breed in the learning process.

2. The value of m, that is, the number of measurements used to compute the numerical variables that physically characterize animals (see Section 2.3). We are assuming that we only know a reduced set of weights that can be used to predict the future weights. Therefore, we consider the first weight of each animal, or the first 2 or 3 or 4. Weights are taken in ascending order of the animals' ages.

The experiments consisted of 5-fold cross-validation (CV) tests. In this kind of test, the dataset is randomly divided into 5 parts, 4 parts are used as training set and the remaining part is used as test set. The learning task is performed on the training set (80% of animals) and the error is computed on the test set (20% of animals). This process is repeated 5 times using each fold as test set. The errors from the 5 test folds are averaged to produce a single score. We establish the folds over the set of animals and then we obtain the number of examples for each animal as explained in Section 3.1. The average number of

Table 5: Number of examples of datasets divided by sex and father's breed (*Avi*: avileña, *Cha*: charolais, *Lim*: limousin, *M*: male, *F*: female)

Father's breed & Sex	Train	Test	Total
Avi_M	25873	6444	32317
Avi_F	1385	345	1730
Cha_M	6584	1674	8258
Cha_F	6169	1572	7741
Lim_M	1888	472	2360
Lim_F	1666	406	2072
	43565	10913	54478

training and test examples used in CV tests is summarized in Table 5.

Some implementation details are presented now. Let us recall that we use a SGD algorithm to find the optimal values for $\boldsymbol{W}$ and $\boldsymbol{V}$ (see Equations 13 and 14). The learning rate in our SGD implementation was defined as:

$$\gamma \leftarrow \frac{1}{(\gamma_s \cdot i) + 1} \tag{25}$$

where i is the iteration number. The program finishes after performing a specified number of iterations.

The values for SGD parameters were automatically selected after a search in the training sets (commonly named grid search) of the best scores. The values checked were:

$$k \in \{2, 10\}$$
$$\gamma_s \in \{10^e : e = -5, \ldots, 0\}$$
$$\nu \in \{0\} \cup \{10^e : e = -8, \ldots, -2\}$$

where k is the dimension of the common space where both animals and ages are mapped, γ_s regulates the decreasing speed of the learning rate and ν is used to weight the regularization.

Recall that test sets were used to assess the performance of the models. So given a test set of n examples, where $lw_1 \ldots lw_n$ are the real weights and $lw'_1 \ldots lw'_n$ are the predicted weights, we compute the accuracy of the model by means of three error measures:

- MSE (Mean Square Error)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (lw'_i - lw_i)^2 \tag{26}$$

- MAE (Mean Absolute Error)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |lw'_i - lw_i| \tag{27}$$

- MAPE (Mean Absolute Percentage Error)

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|lw'_i - lw_i|}{lw_i} * 100 \tag{28}$$

The aim of the experiments is to analyze the variations in performance of the factorization method proposed in this paper depending on the number of measurements, m, and the use, or not, of sex and breed features. In addition, we compare this factorization approach with the alternative regression method. Let us recall that the learning task is aimed at predicting the animals' weights at a any age, using the variables described in Section 3.1.

3.3. Scores and discussion

From the point of view of accuracy, it is clear that the more weights are used in the process of learning, the better results will be obtained. However, breeders are interested in making reliable predictions on the growth of their animals as soon as possible. For example, the sooner it is determined that an animal is not going to achieve certain goals (one of them can be the weight), the sooner it can be discarded for selection purposes. Also, if an animal is in a feedlot, it can be estimated the weight at different ages to decide the best time for slaughter.

Table 6: Effect of the number of weights (m) used to physically characterize an animal for both factorization and regression methods. Error scores are expressed in the form $mean \pm std$

	Factorization			Regression		
m	MSE	MAE(kg)	MAPE(%)	MSE	MAE(kg)	MAPE(%)
1	0.057±0.11	19.9±18.7	5.2±4.7	0.073±0.13	23.6±20.4	6.4±5.3
2	0.038±0.08	16.0±15.6	4.2±3.8	0.053±0.10	19.5±17.8	5.4±5.1
3	0.024±0.06	12.7±12.6	3.5±3.3	0.042±0.08	17.4±15.8	5.1±5.3
4	0.017±0.04	10.8±10.5	3.1±3.2	0.038±0.07	16.8±14.7	5.1±5.2

For this reason, the goal of the first set of experiments is to determinate how the accuracy of our algorithm is influenced by the number of weights, m, used to characterize the physical condition of the animals (see Section 2.3). We used datasets with all animals, with no division by sex or breed. The value of m ranges from 1 to 4, where weights are taken in ascending order of the animals' ages.

Table 6 collects the errors obtained on CV tests. Scores have been computed using the standardized values of the weights. However, MAE scores are expressed in kilos for a better interpretation. As expected, errors decreased as the value of m increases. However, this reduction is much smaller between $m = 3$ and $m = 4$. Therefore, $m = 3$ is a good trade-off between the amount of needed information and the performance of the predictions; it can be established as an appropriate value to make reliable decisions based on such predictions. Moreover, we observe that our approach outperforms linear regression in all cases. This is because, as discussed in Section 2.2, the weight trajectories obtained by regression have the same shape for all animals, while our proposal is able to adapt the trajectories to each animal individually.

So far, we have shown that the model presents good results for all animals put together. We have also separated the animals by sex and father's breed, in order to check how good this general model was for each segment of the animal population. Table 7 shows the scores.

Table 7: Scores computed using the *general model* for each group of animals (with m = 3). *General model* was obtained from a dataset with all animals put together (with no division by sex or father's breed). Error values are expressed in the form *mean±std*. W_AVG is the weighted average per number of examples

Father's breed & Sex	MSE	MAE(kg)	MAPE(%)	#Examples
Avi_M	0.025±0.061	12.4±12.9	3.26±3.16	32317
Avi_F	0.042±0.046	17.1±15.6	5.31±4.58	1730
Cha_M	0.033±0.062	14.2±14.1	3.60±3.38	8258
Cha_F	0.021±0.047	11.8±12.2	3.58±3.71	7741
Lim_M	0.030±0.072	12.9±13.4	3.24±2.98	2360
Lim_F	0.038±0.058	15.8±15.5	4.71±4.31	2072
		W_AVG:	3.48±3.35	

A new set of experiments has been designed to obtain *specific* models for each group of animals. Thus, we obtain 6 models using datasets with the animals divided by sex and by father's breed. We want to compare the accuracy of the *specific* models to the accuracy of the *general* model (when it is applied separately to each group of animals). Obviously, the variables for sex and for father's breed have been eliminated from the examples. In all cases, the value $m = 3$ has been used. Table 8 collects the errors on the *specific* models.

Comparing the scores in Tables 7 and 8 it is observed that accuracy is increased in all cases using *specific models*. The main reason is that the number of examples available for each group of animals is very different. On one hand, this makes the *general model* fit much more to the larger group (Avi_M). On the other hand, we do not have enough examples of the other groups of animals so that the father's breed and sex variables compensate for this over-fit to the Avi_M group. Therefore, with the available data, it is better to learn *specific models* for each group of animals.

Finally, we test whether our method is still better than regression when animals are separated by sex and father's breed. Then, we have calculated the

Table 8: Scores computed using *specific models* for each group of animals (with m = 3). *Specific models* were obtained from datasets of animals divided by sex and father's breed. Error values are expressed in the form *mean±std*. W_AVG is the weighted average per number of examples

Father's breed & Sex	MSE	MAE(kg)	MAPE(%)	#Examples
Avi_M	0.024±0.060	12.3±12.9	3.23±3.17	32317
Avi_F	0.021±0.043	9.8±9.4	3.07±2.75	1730
Cha_M	0.021±0.061	12.6±13.6	3.18±3.25	8258
Cha_F	0.020±0.047	8.7±8.6	2.64±2.62	7741
Lim_M	0.023±0.070	12.8±13.9	3.17±3.08	2360
Lim_F	0.021±0.055	8.8±9.3	2.63±2.58	2072
		W_AVG:	3.11±3.06	

scores by means of regression for each group of animals separately. They are shown in Table 9.

As we can observe comparing Tables 8 and 9, our method also achieves better scores than regression when animals are separated into groups by sex and father's breed. Let us emphasize again the main reason for this advantage: the factorization method proposed in this paper is able to adapt the weight trajectories to each animal *individually*, while the regression method is not, all the curves are parallel. As an example, Figure 2 contains the weight trajectories of 4 bulls randomly selected from *Avi_M* group. These growth curves have been obtained with our factorization method. Notice that each animal has a customized curve.

3.4. Visualization in the Euclidean space

In addition to predicting weights with a very good performance, the factorization method allows projecting the information of animals in a common Euclidean space. Remember that this common space summarizes all the information of each object (animal or age) in a vector of k components. Visualizing information when the Euclidean space has up to 3 dimensions (up to $k = 3$) can

Table 9: Scores computed using regression for animals divided by sex and father's breed (with m = 3). Error values are expressed in the form $mean \pm std$. W_AVG is the weighted average per number of examples

Father's breed & Sex	MSE	MAE(kg)	MAPE(%)	#Examples
Avi_M	0.034±0.069	15.6±14.5	4.39±4.52	32317
Avi_F	0.029±0.051	12.2±10.3	4.04±3.68	1730
Cha_M	0.033±0.069	17.0±15.7	4.59±4.78	8258
Cha_F	0.036±0.068	12.0±10.9	3.87±3.93	7741
Lim_M	0.027±0.056	15.3±14.1	4.02±3.77	2360
Lim_F	0.034±0.067	12.1±10.7	3.78±3.37	2072
		W_AVG:	4.30±4.37	

be very useful to interpret the results. Moreover, clustering techniques can be used in order to find groups of animals with similar characteristics according to a specific trait.

Available data on the group of animals AVI_M were used to create Figure 3. We have learned a model applying our factorization algorithm with $k = 2$. Then, we represented in the Euclidean space of animals the best and the worst deciles according to their average daily gain (ADG).

Two different regions can be clearly distinguished. The region on the left contains animals with poor ADG while the region on the right-hand contains animals with good ADG. Obviously, there are many animals that are in an intermediate zone. Nevertheless, this observation is very important because it confirms that some animals can be separated according to some of their traits, like the average daily gain. In this case, we could identify animals that will gain very little weight and those that will gain a lot of weight during their stay in the feedlot. Based on this information, breeders might decide to slaughter each animal at different ages according to their economic performance criteria. In the same way, clusters based on other characteristics could be considered, such as stockbreeder, feedlot or father's breed.
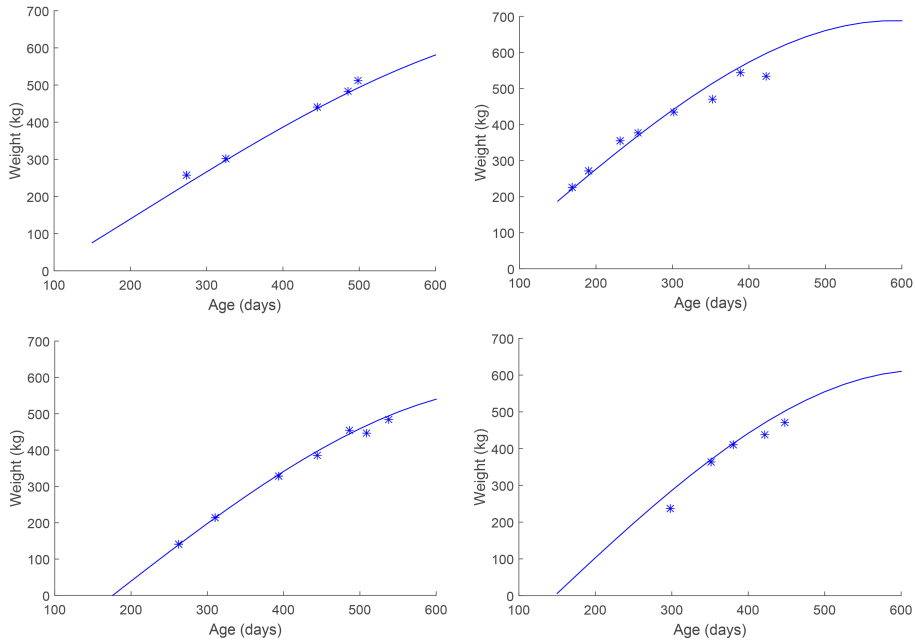
Figure 2: Weights of 4 randomly selected animals from Avi_M and their corresponding learned growth trajectory. Our factorization method results in customized trajectories for each animal

4. Conclusions

We have presented a factorization approach for learning the growth trajectories of beef cattle during their stay in a feedlot. These trajectories are adapted to the specific data of each animal, which yields an improved precision with respect to a linear estimate. This means that the estimation of the individualized evolution of the weights requires a nonlinear component in which the interactions between all available descriptors about animal and age associated with a weight are considered. In addition, the representation of animals in a common space of reduced dimension allows establishing clusters of animals with similar characteristics.
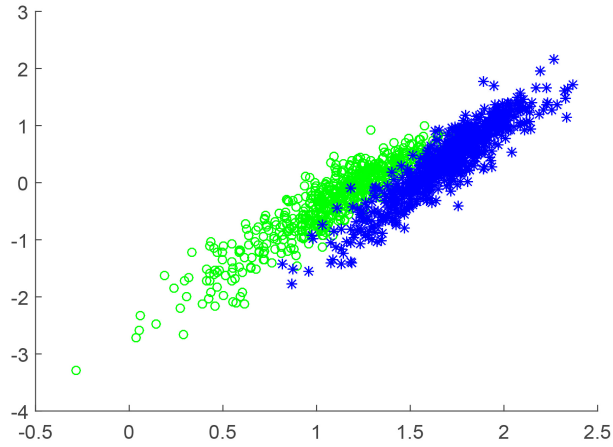
Figure 3: Visualization in a 2-dimensional Euclidean space (k=2) of some animals of the group Avi_M. We use (*) to represent animals with good average daily gain (ADG) and (o) to represent animals with poor ADG

Acknowledgements

References

Albertí, P., Panea, B., Sañudo, C., Olleta, J.L., Ripoll, G., Ertbjerg, P., Christensen, M., Gigli, S., Failla, S., Concetti, S., Hocquette, J.F., Jailler, R., Rudel, S., Renand, G., Nute, G.R., Richardson, R.I., Williams, J.L., 2008. Live weight, body size and carcass characteristics of young bulls of fifteen European breeds, Livestock Science 114 (1), 19–30.

Alonso, J., Bahamonde, A., Villa, A., Castañón, Á. R., 2007. Morphological

assessment of beef catle according to carcass value. Livestock Science 107, 265–273.

Alonso, J., Castañón, Á. R, Bahamonde, A., 2013. Support Vector Regression to predict carcass weight in beef cattle in advance of the slaughter. Computers and Electronics in Agriculture 91, 116–120.

Alonso, J., Villa, A., Bahamonde, A., 2015. Improved estimation of bovine weight trajectories using Support Vector Machine Classification. Computers and Electronics in Agriculture 110, 36–41.

Arango, J.A., Cundiff, L.V., Van Vleck, L.D., 2004. Covariance functions and random regression models for cow weight in beef cattle. Journal of Animal Science 82 (1), 54–67.

Bottou, L., 2010. Large-Scale Machine Learning with Stochastic Gradient Descent. Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics, 177–186.

Coopman, F., De Smet, S., Laevens, H., Van Zeveren, A., Duchateau, L., 2009. Live weight assessment based on easily accessible morphometric characteristics in the double-muscled Belgian Blue beef breed. Livestock Science 125 (2-3), 318–322.

Díez, J., Bahamonde, A., Alonso, J., López, S., del Coz, J. J., Quevedo, J., Ranilla, J., Luaces, O., Alvarez, I., Royo, L., Goyache, F., 2003. Artificial intelligence techniques point out differences in classification performance between light and standard bovine carcasses. Meat Science, 64, 249–258.

Díez, J., del Coz, J.J., Bahamonde, A., Sañudo, C., Olleta, J.L., Macie, S., Campo, M.M., Panea, B., Albertí, P., 2006. Identifying market segments in beef: Breed, slaughter weight and ageing time implications, Meat Science 74 (4), 667–675.

Enevoldsen, C. and Kristensen. T., 1997. Estimation of body weight from body size measurements and body condition scores in dairy cows, Journal of Dairy Science 80 (9), 1988–1995.

Freetly, H. C., Kuehn, L. A. and Cundiff, L. V., 2011. Growth curves of crossbred cows sired by Hereford, Angus, Belgian Blue, Brahman, Boran, and Tuli bulls, and the fraction of mature body weight and height at puberty. Journal of Animal Science 89 (8) 2373–2379.

Jaffrézic, F., Pletcher, S.D., 2000. Statistical Models for Estimating the Genetic Basis of Repeated Measures and Other Function-Valued Traits. Genetics 156, 913–922.

Ji, Y. and Eisenstein, J., 2014. Representation Learning for Text-level Discourse Parsing. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL (1), 13–24.

Kanagal, B., Ahmed, A., Pandey, S., Josifovski, V., Garcia-Pueyo, L. and Yuan, J., 2013. Focused matrix factorization for audience selection in display advertising. Data Engineering (ICDE), IEEE 29th International Conference on, Brisbane 2013, 386–397.

Koren, Y., Bell, R., Volinsky, C., 2009. Matrix factorization techniques for recommender systems. Computer 42 (8), 30 –37.

Khurana, P., Bhattacharjee, P., Majumdar, A., 2015. Matrix factorization from non-linear projections: application in estimating T2 maps from few echoes Magnetic Resonance Imaging 33 (7), 927–931.

Luaces, O., Díez, J., Alonso-Betanzos, A., Troncoso, A., Bahamonde, A., 2015. A factorization approach to evaluate open-response assignments in MOOCs using preference learning on peer assessments. Knowledge-Based Systems 85, 322–328.

Luaces, O., Díez, J., Joachims, T., Bahamonde, A., 2015. Mapping preferences into Euclidean space. Expert Systems with Applications, Volume 42 (22), 8588–8596.

Moon, K.R., Delouille, V., Li, J., De Visscher, R., Watson, F., Hero III, A., 2016. Image patch analysis of sunspots and active regions - II. Clustering via matrix factorization. J. Space Weather Space Clim. 6 A3.

Stajnko, D., Brus, M., Hocevar, M., 2008. Estimation of bull live weight through thermographically measured body dimensions. Computers and Electronics in Agriculture 61 (2), 233–240.

Tasdemir, S., Urkmez, A., Inal, S., 2011. Determination of body measurements on the holstein cows using digital image analysis and estimation of live weight with regression analysis. Computers and Electronics in Agriculture 76, 189–197.

Wang, S., Pedrycz, W., Zhu, Q., Zhu, W., 2015. Subspace learning for unsupervised feature selection via matrix factorization. Pattern Recognition 48 (1), 10–19.

West, G. B., Brown, J. H., Enquist, B. J., 2001. A general model for ontogenetic growth. Nature 413, 628–631.

Zhang, Y., Chen, M., Huang, D., Wu, D., Li, Y., 2016. iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. Future Generation Computer Systems. Available online 12 January 2016, http://dx.doi.org/10.1016/j.future.2015.12.001.